

Enhanced Efficiency of Mapping Distribution Protocols in Scalable Routing and Addressing Architectures

Kotikalapudi Sriram, Young-Tak Kim, and Doug Montgomery

Abstract

In this paper, we present a discussion of some architectural principles pertaining to mapping distribution protocols that are used in solutions for scalability of future Internet routing tables. The efficiency of a mapping distribution protocol in terms of response time and the volume of traffic load it generates are important considerations. We consider how Egress Tunnel Routers (ETRs) can perform aggregation of end-point ID (EID) address space belonging to their downstream delivery networks, in spite of migration/re-homing of some subprefixes to other ETRs. This aggregation may be useful for reducing the processing load and memory consumption associated with mapping messages, especially in some resource-constrained components of the mapping distribution system. Consequently, the proposed methods can also help improve response time (e.g., first packet delay). Some interesting architectural issues, their potential solutions and trade-offs are discussed. The overarching goal is to expose and discuss some subtleties in design considerations for mapping distribution and management.

Index Terms

Future Internet Architecture, Scalable Internet, Scalable Routing and Addressing, Mapping Distribution Protocol, Locator and ID Separation, LISP, Aggregation

I. INTRODUCTION

The problem of scalability of today's Internet routing system has been discussed in recent IETF reports[1][2]. Several proposals aimed at solving this problem are under study in IRTFs Routing Research Group (RRG)[3]-[11]. In this paper, we present some architectural principles pertaining to the mapping

K. Sriram and D. Montgomery are with the National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA (Email: {ksriram, dougm}@nist.gov). Y.T. Kim was a guest researcher at NIST and is currently with Yeungnam University, Gyeongsan, Korea (Email: ytkim@yu.ac.kr).

distribution protocols (MDPs) [9]-[11], especially applicable to map-and-encap class of solutions. Aforementioned architectural principles enhance the efficiency of MDPs in terms of (1) Better utilization of resources (e.g., processing and memory) at Ingress Tunnel Routers (ITRs) and mapping servers, and consequently, (2) Reduction of response time (e.g., first packet delay). We consider how Egress Tunnel Routers (ETRs) can perform aggregation of end-point ID (EID) address space belonging to their downstream delivery networks, in spite of migration/re-homing of some subprefixes to other ETRs. We discuss incorporation of an exception message as part of the map announcement message to indicate that portions of the ID space (some small number of more specific prefixes or subprefixes) under a less specific prefix have moved to or reside at different ETRs. This aggregation may be useful for reducing the processing load and memory consumption associated with map messages, especially at some resource-constrained ITRs and subsystems of the mapping distribution system. We also consider another architectural concept where the ETRs are organized in a hierarchical manner for the potential benefit of aggregation of their EID address spaces. The aforesaid architectural principles are described and discussed in detail in Sections III and IV. Conclusions and possible directions for extending this study are stated in Section V.

II. ARCHITECTURAL FRAMEWORK FOR ID TO LOCATOR MAPPING

We feel that the architectural principles examined here are generally applicable to several of the proposals being discussed in the RRG (summarized in [3]), and their associated mapping distribution protocols. An example of map-and-encap solutions is the Locator ID Separation Protocol (LISP) protocol[4], and here we use LISP-like high level architecture to describe our proposal for enhancing the efficiency of mapping distribution and management. The specific MDP associated with LISP is known as LISP+ALT[9]. To assist in this discussion, we start with the high level architecture of a map-and-encap approach as illustrated in Fig. 1. This helps anchor the discussion of the principles of the mapping distribution protocol to an architectural framework; the specific architecture can, however, vary and the ideas presented here are generally relevant to any of the proposals that are currently being reviewed in RRG. As shown in Fig. 1, the Egress Tunnel Routers (ETRs) generate map messages to inform the ID-Locator Mapping (ILM) servers of end-point ID prefixes or delivery-network address ranges that can be reached through them. The ILMs are repositories for complete mapping information, while the ILM-Regional (ILM-R) servers can contain partial and/or regionally relevant mapping information. The ETRs can push ID-to-locator mapping information pertaining to the delivery networks under their purview to an ILM-R or an ILM they are associated with. The ILM-Rs push the mapping information received from ETRs to the

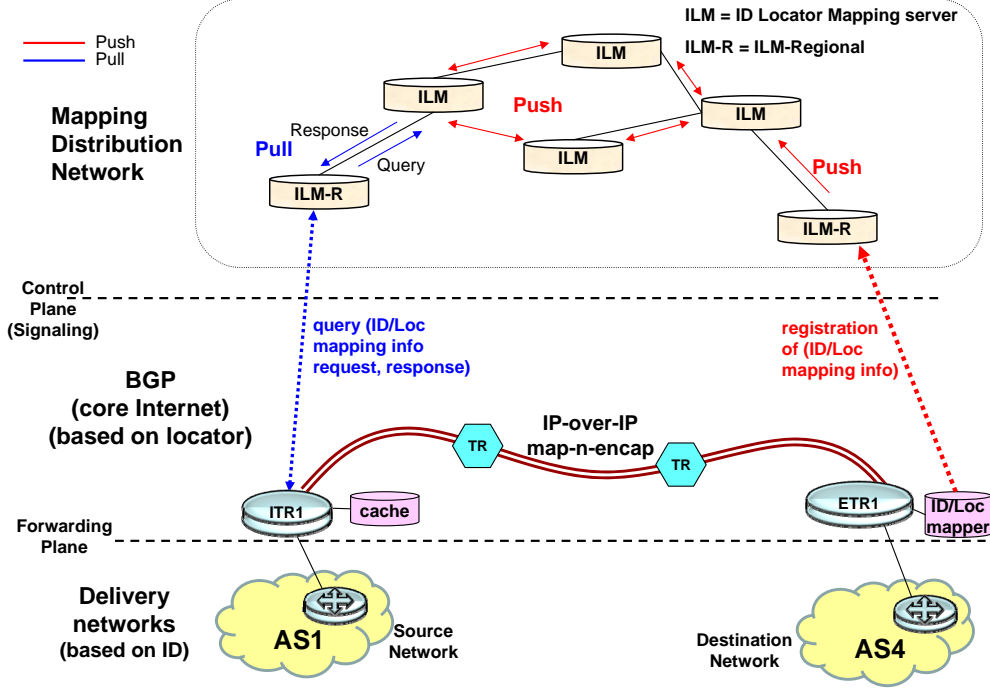


Fig. 1. An example of map-and-encap and mapping distribution architecture.

ILMs, and the ILMs update each other by pushing the mapping updates as they are received. When an Ingress Tunnel Router (ITR) does not have mapping information for an EID, it initiates pull requests (query/response) to the ILM-R that it has access to. The ILM-R can then pull the relevant mapping information from an ILM (if necessary) and respond back to the ITR.

III. MAPPING DISTRIBUTION OF SUBPREFIXES SPREAD ACROSS MULTIPLE ETRs

With the help of Fig. 2, it is illustrated that while a large endpoint address space contained in a prefix may be mostly associated with the delivery networks served by an ETR, some fragments (subprefixes) of that address space may be located elsewhere at different ETRs. Let $a/20$ denote a prefix that is conceptually viewed as composed of 16 subprefixes of $/24$ size that are denoted as $a_0/24, a_2/24, \dots, a_{15}/24$. In Fig. 2, for example, $a/20$ is mostly at ETR1, while only two of its subprefixes $a_7/24$ and $a_{14}/24$ are elsewhere at ETR3 and ETR2, respectively. From the point of view of efficiency of the mapping distribution protocol, it may be beneficial for ETR1 to announce a map for the entire space $a/20$ (rather than fragment it into a multitude of more-specific prefixes), and provide the necessary exceptions in the map information. Thus the map response could be in the form of *Map*:($a/20$, ETR1; *Exceptions*: $a_7/24, a_{14}/24$). In addition, ETR2 and ETR3 announce the maps for $a_{14}/24$ and $a_7/24$ respectively, and so the ILMs know where the

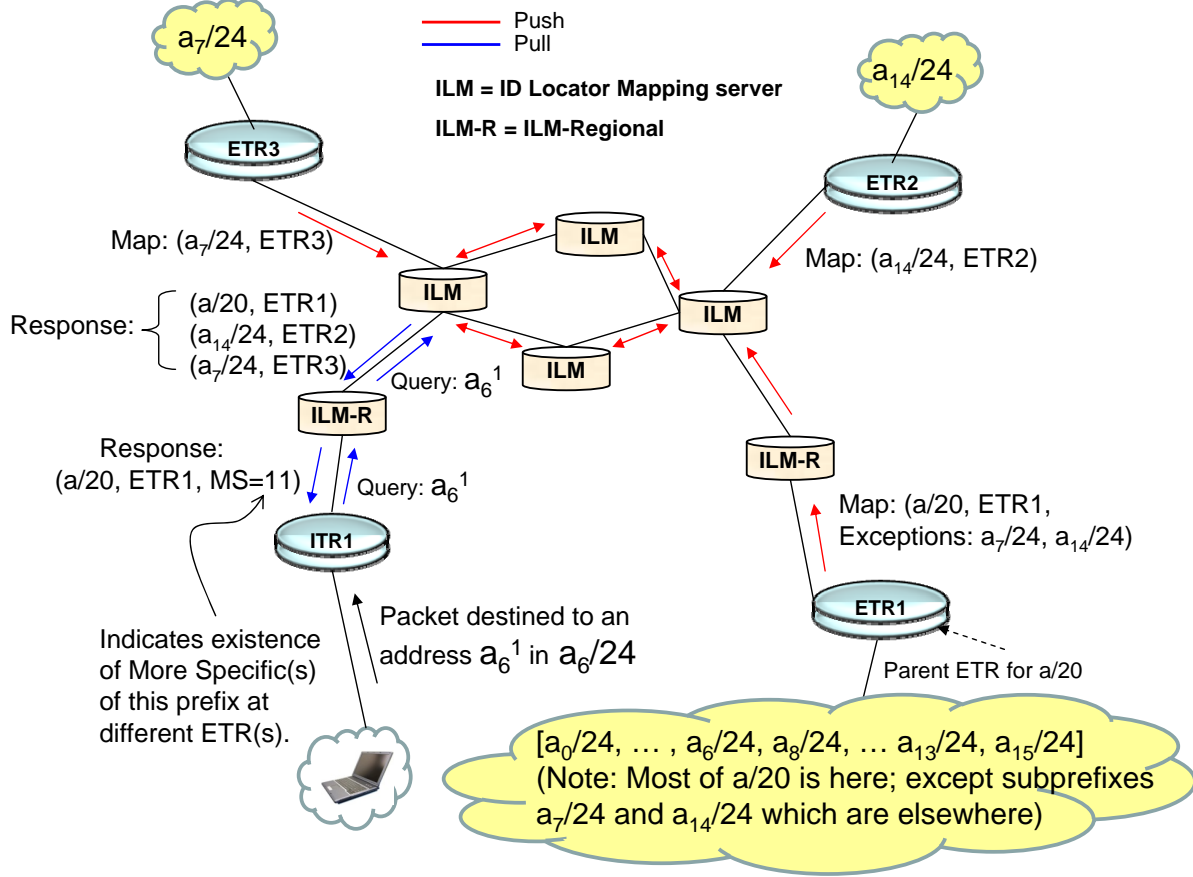


Fig. 2. Illustration of endpoint ID aggregation and exception concepts.

exception endpoint ID addresses are located. The details of how the map responses are structured and communicated will be described shortly.

To support the above assertions with numerical data, we provide an actual example and some statistics regarding holes in prefixes based on Internet Corporation for Assigned Names and Numbers (ICANN) reports [12] and Routeviews trace data [13]. From ICANN BGP reports [12], we observe, as an example, that prefixes 129.6.0.0/17 and 129.6.0.128/17 are originated from AS49, and additionally 129.6.112.0/24 is originated from AS10886. Thus we notice that there is a prefix-hole (or exception) in that 129.6.112.0/24 is originated from AS10886, while the rest of the prefix 129.6.0.0/17 (in fact the rest of 129.6.0.0/16) is originated from AS49. Fig. 3 illustrates how this kind of prefix-hole results in a substantial increase in number of EID-to-locator map entries for map-and-encap schemes. In Fig. 4, measurements based on Routeviews trace data [13] are reported with a plot of the number of prefix-holes vs. the corresponding mask-length. The y-axis in Fig. 4 is essentially the number of subprefixes of length x (x being value on

Aggregate: 129.6.0.0/16	
Announced in BGP-4:	
129.6.0.0/17	Origin: AS49
129.6.0.128/17	Origin: AS49
129.6.112.0/24	Origin: AS10886
EID to Locator (ETR) Mapping:	
EID:	ETR (equivalent)
129.6.0.0/18	ETR49
129.6.64.0/19	ETR49
129.6.96.0/20	ETR49
129.6.112.0/24	ETR10886
129.6.113.0/24	ETR49
129.6.114.0/23	ETR49
129.6.116.0/22	ETR49
129.6.120.0/21	ETR49
129.6.128.0/17	ETR49

Fig. 3. Impact of a prefix-hole on multiplication of EID-to-locator map entries.

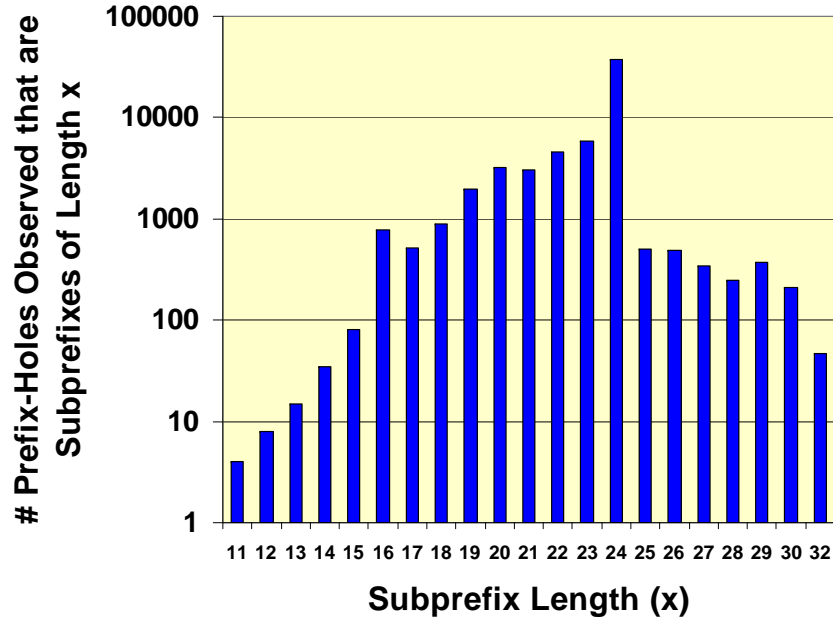


Fig. 4. Measurement of numbers of prefix-holes of various mask-lengths (from Routeviews trace data).

the x-axis) such that each has a less-specific announced with a different origin AS as compared to that of the subprefix. Many of these situations, if not all, would result in holes in the EID-to-locator maps. The trace data used in Fig. 4 consisted of 304,723 RIB entries from which a total of 60,988 were identified as likely prefix-holes. Based on this type of empirical evidence, we feel that there is a strong need to address the map-proliferation problem with suitable enhancements to the mapping distribution protocol.

Now the interesting question is in what ways does the mapping distribution system communicate mapping information to the ITRs while coping with EID prefix-holes (or exceptions)? This issue is illustrated in the lower left-hand portion of Fig. 2. The sending host initiates a packet destined for an address a_6^1 , which is in $a_6/24$ and hence in the normal portion of $a/20$ and not in the exception portion. In this case, we assume that the ITR1 does not have the map information, so it sends a query to its ILM-R. Assuming that the ILM-R does not have the map information either, it will send a query to an ILM it is connected with. The ILM can then send a response to the ILM-R, and this response can contain the information regarding matching prefix ($a/20$, ETR1) as well as the maps for the exceptions or the subprefixes that are elsewhere, namely, ($a_{14}/24$, ETR2) and ($a_7/24$, ETR3). Now a question arises as to which of the following approaches would be the best choice:

- 1) ILM-R provides the complete mapping information for $a/20$ to ITR1 including all the maps for the relevant exception subprefixes.
- 2) ILM-R provides only the directly relevant map to ITR1 which in this case is ($a/20$, ETR1).
- 3) The mapping information transaction between ILM-R and ITR1 can dynamically use approach 1 or approach 2 above depending on the context (further explanation of this is provided below).

In the first approach, the advantage is that ITR1 would have the complete mapping for $a/20$ (including exception subprefixes), and it would not have to generate queries for subsequent first packets that are destined to any address in $a/20$, including $a_7/24$ and $a_{14}/24$. This would be true as long as ITR1 holds the received mapping information in its memory. However, the disadvantage is that if there is a significant number of exception subprefixes (in the example in consideration there are only two but in general it can be many more), then the very first packet destined for $a/20$ will experience a long delay, and also the processors at ITR1 and ILM-R can experience overload. In addition, the memory usage at ITR1 can be very inefficient as well.

The advantage of the second approach above is that the ILM-R does not overload resources at the ITR both in terms of processing and memory usage. This will help avoid resource exhaustion at the ITRs and thus provide better response time in handing mapping queries for the packets received from hosts. However, some care must be exercised here. ITR1 might save the mapping information for $a/20$ and reuse it. So it should be at least made aware that possibly there are subprefixes under $a/20$ that are at other ETRs. This can be taken care of by incorporating an More Specific (MS) indicator in the map response sent to ITR1 as shown in Fig. 2. This map response is of the form *Map*:($a/20$, ETR1, MS=11). The MS indicator is set to a binary value 11 to indicate to ITR1 that not all addresses in $a/20$ map to

Prefix	ETR	MS	K	SN	NML
--------	-----	----	---	----	-----

MS = More Specific indicator

K = # Maps to follow

SN = Sequence Number ($\leq K + 1$)

NML = Next longer Mask-Length

Fig. 5. Illustration of a format for map response.

ETR1, and accordingly ITR1 will re-enquire next time there is a packet destined for another address in $a/20$ (other than a_6^1). One can go into the details of this methodology and improve it further; the detailed algorithm is described in Subsection III-A. The key idea here is that aggregation is beneficial, and subprefix exceptions must be handled with additional messages or indicators in the map.

The third approach above seeks adaptability to use either the first approach or the second depending on the context. Here a threshold (or limiting value), say H , can be applied to the number of map responses (map for the parent prefix plus all maps for the relevant exceptions subprefixes). If this threshold is not exceeded, then the first approach would be used, else, the second approach would be used. This parameter can also be tuned administratively or dynamically (depending on the state in terms of resource availability at the ILM and/or the ITR).

A. Details of the Map Response Algorithm

In Fig. 5, a candidate format is illustrated for the map response sent to an ITR. The More Specific (MS) indicator is a two-bit field, and is used to specify that there are additional maps available for more specific prefixes; the exact details of usage of the MS field will be described shortly. The number of additional more specific maps that would be communicated (following the initial map response) is denoted by the variable K . The map response also carries a sequence number (SN); when SN is set to a value i , it indicates that the present map response is the i^{th} out of a total of $K+1$ map responses being sent (in response to a map query). The NML field carries the value of next longer mask-length (NML), and its usage will be described shortly as well.

The details of map response generation algorithm can be described by explaining the usage of the format parameters that are shown in Fig. 5. The usage and interpretation of these parameters, namely MS, K, SN, and NML, is explained in Table I. There are four cases the algorithm should encompass as enumerated below (also refer to Table I).

TABLE I
PARAMETERS AND THEIR INTERPRETATION IN THE MAP RESPONSE ALGORITHM

Case #	More Specific Indicator (MS)	No. of Maps (K)	Sequence No. (SN)	Next Longer Mask-Length (NML)	Interpretation
Case 1	00	1	1	Don't Care	Map response has no exceptions.
Case 2	01	$k + 1$	$i \ (1 \leq i \leq k + 1)$	Don't Care	Map response has exceptions; Additional k map responses for the exception subnets will follow automatically.
Case 3	10	$k + 1$	$i \ (1 \leq i \leq k + 1)$	Don't Care	Map response has exceptions; Additional k map responses for the exception subnets follow automatically but the ETR information for one or more specific subnets is "Don't Care" (because the specific subnet is further split, i.e., multiple sub-subnets exist in said subnet with different ETRs).
Case 4	11	1	1	m	Map response has exceptions; Additional map responses are not provided because # subnets (at next longer mask-length value) exceeds a threshold (H); the next longer mask-length value, $NML = m$, is provided here.

- 1) **There is only one unique map response:** In responses to the map request, the ILM has only one unique {prefix, ETR} mapping response. There are no exceptions subprefixes in this case. The parameters in the map response are set to: MS = 00, K = 1, SN = 1, and NLM = Don't Care (i.e., some default code to signify Don't Care).
- 2) **All available exception messages communicated:** In responses to the map request, the ILM has multiple {prefix, ETR} mapping responses to notify the ITR about the immediately usable {prefix, ETR} mapping as well as mappings for subprefixes within that prefix. In this case, all exception subprefix mappings are notified. The parameters in the map response are set to: MS = 01, K is set to one plus # subprefixes, SN varies from 1 to K, and NLM = Don't Care.
- 3) **Subnets are further split into sub-subnets:** This case is similar to Case 2 above, but here the ETR information for one or more specific subnets is denoted as "Don't Care" in the map responses. This is because the specific subnet is further split, i.e., multiple sub-subnets exist in said subnet with different ETRs. In this case, the parameters in the map response are set to: MS = 10, K is set to one plus # subprefixes, SN varies from 1 to K, and NLM = Don't Care.
The usage scenario for this case can be explained with an example. Consider a scenario when a

/24 subprefix is split from a corporate /20 prefix and company's mobile devices are allocated IP addresses from that one /24. The /20 is homed to ETR1 (except the /24) while the mobile devices are homed to many different ETRs elsewhere. The traffic from the ITR is almost entirely destined for addresses in the /20 other than the mobile-designated /24 part. So the ITR benefits by knowing that if it ever has traffic for an address in the mobile-designated /24, then it needs to re-query. Hence, the ITR having knowledge of just the exception subprefix - without knowing about its ETR - works just fine as long as the traffic is destined for other addresses in the /20.

- 4) **Only the next longer mask-length (NML) for exception subnets is communicated:** In this case, the number of exception subnets at the next longer mask-length value (relative to prefix mask-length for the relevant $\{p, \text{origin}\}$ map-response that is being communicated) is too high and exceeds a threshold (H). So only the NML ($=m$) information is provided along with the immediately relevant $\{p, \text{origin}\}$ map-response. In this case, the parameters in the map response are set to: $MS = 11$, $K = 1$, $SN = 1$, and $NLM = m$.

The usefulness of this case can be explained with an example scenario. Consider a scenario when Org-A has a /16 prefix; bulk of it resides at the headquarters (ETR1); several /24s subprefixes (within the /16 prefix) are homed to other ETRs elsewhere; majority of the traffic for the /16 is destined to a specific /24 subnet located at the headquarters (ETR1). Say the very first packet of the very first session at an ITR already fetched a map for said specific /24 subnet. Most subsequent packets are very likely to be destined for addresses in this /24 EID subprefix. Thus, as long as subsequent packets from that ITR are destined for addresses for which the destination EID address matches up to m -bits (noting that $NML = m$), the ITR need not generate any new map queries.

IV. MAPPING DISTRIBUTION FOR SCENARIOS WITH HIERARCHY OF ETRs AND MULTI-HOMING

Now we refer to Fig. 6 to highlight another architectural concept related to mapping management. Here we consider the possibility that ETRs may be organized in a hierarchical manner. For instance ETR7 in Fig. 6 is higher in the loose hierarchy relative to ETR1, ETR2, and ETR3, and like-wise ETR8 is higher relative to ETR4, ETR5, and ETR6. For instance, ETRs 1 through 3 can relegate locator role to ETR7 for their EID address space. In essence, they can allow ETR7 to act as the locator for the delivery networks in their purview. ETR7 keeps a local mapping table for mapping the appropriate EID address space to specific ETRs that are hierarchically associated with it in the level below. In this situation, ETR7 can perform EID address space aggregation across ETRs 1 through 3 and can also include its own immediate EID address space for the purpose of that aggregation. Thus in the example of Fig. 6, ETR7 can aggregate

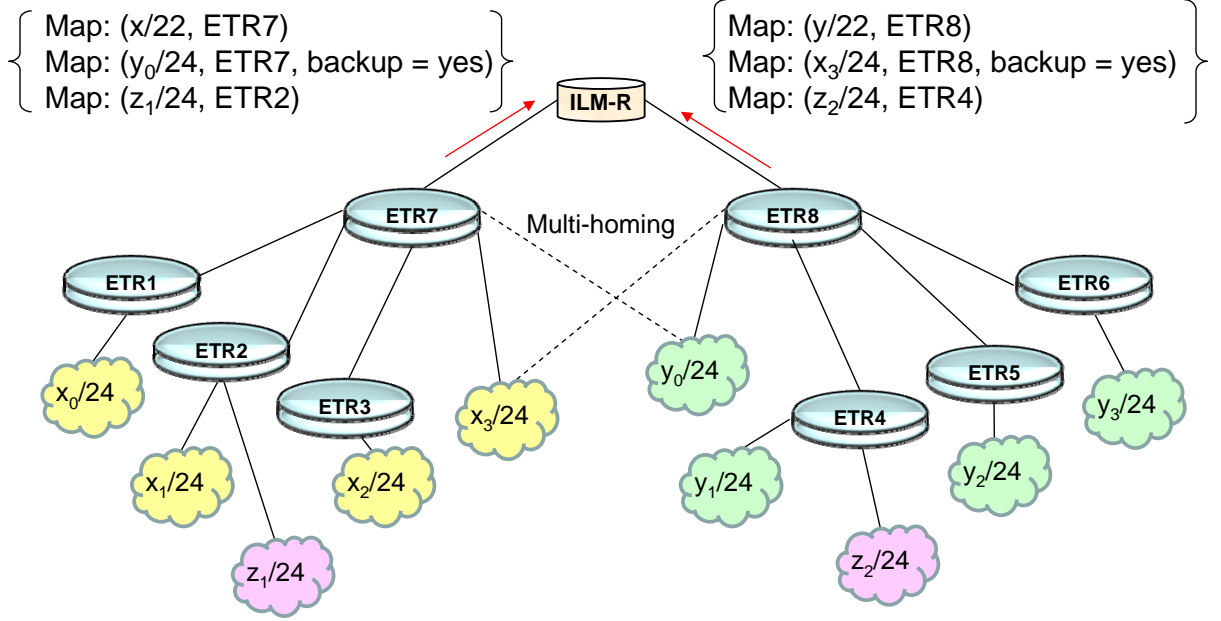


Fig. 6. Illustration of endpoint ID aggregation in the presence of multihoming.

$x_0/24$, $x_1/24$, $x_2/24$, $x_3/24$ into a map message of the form $Map:(x/22, ETR7)$ to inform ILM-R. Similarly, ETR8 can aggregate $y_0/24$, $y_1/24$, $y_2/24$, $y_3/24$ into a map message of the form $Map:(y/22, ETR8)$. This architectural principle again lessens the possibility of cluttering the mapping distribution system with excessive map messages or entries. It may be noted that ETR7 and ETR8 may or may not hide the ETRs below in the hierarchy. This is because there may be prefixes (subnets) such as $z_1/24$ at ETR2 and $z_2/24$ at ETR4, which benefit by simply announcing $Map:(z_1/24, ETR2)$ and $Map:(z_2/24, ETR4)$, respectively. The idea is that when an ETR at the higher level is not able to aggregate EID prefixes of some ETRs below in the hierarchy, the routing locator may as well be the actual ETR where the delivery network resides.

Another architectural consideration for mapping distribution and management arises when some delivery networks (i.e., EID address space prefixes) are multi-homed to different ETRs. Examples of this in Fig. 6 are $x_3/24$ and $y_0/24$ which are each multi-homed to ETR7 and ETR8. In such case, ETR7 distributes a map message for $y_0/24$ which is a more specific of $y/22$ for which ETR8 distributes a map message. The map message for such subprefix multi-homing situations may incorporate a field as shown in Fig. 6 to indicate that the ETR in the map provides a backup or lower priority path. But it is up to the ITRs as to how they make use of the backup/priority information in such cases. They

may go by the priority information in the map response or they may prefer more specifics in any case. Recommendations for those actions or choices can be made in the future based on performance and operational considerations. In summary, the takeaway from Fig. 6 is that additional EID address space aggregation opportunities exist if there is at least a loose hierarchical structure to the ETRs. However, some minor complexities can also arise in some multi-homing cases, which can be handled by suitable best common practice (BCP) recommendations. These recommendations can be established in the future based on network operators' goals and preferences.

The illustration in Fig. 6 also brings to mind a question about whether it would be a good design choice to recursively perform map-and-encap routing via a hierarchy of ETRs. This question has been brought up for discussion in some other proposals also that have been shared in the RRG. The hierarchical organization of ETRs and delivery networks potentially helps in the future growth and scalability of mapping distribution protocol as well. With recursive map-and-encap, some of the mapping distribution and management functionality will remain local to topologically neighboring delivery networks which are hierarchically underneath ETRs that reside in transit networks and serve as gateways.

V. CONCLUSION

We have discussed various architectural questions, and offered various solutions related to the mapping distribution and management in map-and-encap class of solutions for Internet routing and addressing scalability. We have proposed a methodology for aggregation possibilities for the EID address space associated with delivery networks that are largely homed at the same ETR. We introduced special considerations such as exceptions in the sense that some subprefixes may be located away at different ETRs other than the ETR that is home to bulk of a covering less specific prefix. We also discussed the notion of a loose hierarchy of ETRs with the potential benefit of aggregation of their EID address spaces. The purpose of this paper is to help generate some discussion of these issues. The overarching goal is to expose and discuss some subtleties in design considerations for mapping distribution and management. Hope is that early awareness of these issues may result in anticipating and averting some architectural roadblocks in the future Internet. The end goal is better efficiency and performance for the mapping distribution protocol.

ACKNOWLEDGMENT

The authors would like to thank Lixia Zhang, Dino Farinacci, Robin Whittle, Christian Vogt, and Brian Carpenter for their critique and comments on earlier drafts of this work. We would also like to

thank our NIST colleagues Patrick Gleichmann, Okhee Kim, Oliver Borchert, and Rick Kuhn for their comments and suggestions. This research was supported by the Department of Homeland Security under the Secure Protocols for the Routing Infrastructure (SPRI) program and the NIST Information Technology Laboratory Cyber and Network Security Program.

REFERENCES

- [1] D. Meyer, L. Zhang, and K. Fall, "Report from the IAB Workshop on Routing and Addressing," IETF Internet Draft: draft-iab-raws-report-00.txt, December 2006.
- [2] T. Narten, "On the Scalability of Internet Routing," IETF Internet Draft: draft-narten-radir-problem-statement-05, February 2010. <http://tools.ietf.org/html/draft-narten-radir-problem-statement-05>
- [3] T. Li and L. Zhang, "Recommendation for a Routing Architecture," IETF Internet-Draft: draft-irtf-rrg-recommendation-04, January 2010. <http://tools.ietf.org/html/draft-irtf-rrg-recommendation-04>
- [4] D. Farinacci, V. Fuller, D. Meyer, and D. Lewis, "Locator/ID Separation Protocol (LISP)," IETF Internet Draft: draft-ietf-lisp-06, January 2010. <http://tools.ietf.org/html/draft-ietf-lisp-06>
- [5] R. Whittle, "Ivip (Internet Vastly Improved Plumbing) Architecture," IETF Internet Draft: draft-whittle-ivip-arch-03, January 2010. <http://tools.ietf.org/pdf/draft-whittle-ivip-arch-03>
- [6] C. Vogt, "Six/One: A Solution for Routing and Addressing in IPv6," IETF Internet Draft: draft-vogt-rrg-six-one-02, October 2009. <http://tools.ietf.org/html/draft-vogt-rrg-six-one-02>
- [7] C. Vogt, "Six/One Router: A Scalable and Backwards-Compatible Solution for Provider-Independent Addressing," ACM International Workshop on Mobility in the Evolving Internet Architecture (MobiArch), Seattle, WA, USA, August 2008, pp. 13-17.
- [8] B. Wang, L. Wang, X. Zhao, Y. Liu and L. Zhang, "FIB Aggregation," IETF Internet Draft, draft-zhang-fibaggregation-02, IETF 76, November 2009. <http://tools.ietf.org/html/draft-zhang-fibaggregation-02>
- [9] D. Farinacci, V. Fuller, D. Meyer, and D. Lewis, "LISP Alternative Topology (LISP+ALT)," IETF Internet Draft: draft-ietf-lisp-alt-02, January 2010. <http://tools.ietf.org/html/draft-ietf-lisp-alt-02>
- [10] R. Whittle, "Ivip Mapping Database Fast Push," IETF Internet Draft: draft-whittle-ivip-db-fast-push-03, January 2010, <http://tools.ietf.org/html/draft-whittle-ivip-db-fast-push-03>
- [11] L. Iannone and O. Bonaventure, "On the Cost of Caching Locator/ID Mappings," 3rd Annual CoNEXT Conference, December 2007. <http://inl.info.ucl.ac.be/system/files/Conext-2007-CRV-UCL-v2-Clean.pdf>
- [12] BGP Reports, Published by ICANN Research. <http://stats.research.icann.org/bgp/>
- [13] Route Views Project, University of Oregon. <http://www.routeviews.org/>